# The Usefulness of the Double Entry Constraint for Predicting Earnings

Dr Ehsan Khansalar[1] and Dr Eilnaz Kashefi-Pour[2]

*[1] Kingston Hill Campus, Kingston University, London, KT2 7LB (Corresponding author)*

*Mobile: +447528593389 Email: E.khansalar@kingston.ac.uk*

*[2] Birmingham Business School, University of Birmingham*

*Email: E.KashefiPour@bham.ac.uk*

# The Usefulness of the Double Entry Constraint for Predicting Earnings

**Abstract** In the absence of an income statement, earnings can be calculated as cash flow from operating activities (CFO) plus accruals, rather than being stated as the difference between income statement revenues and expenses. Following the study by Christodoulou and McLeay (2014), this paper uses a system of structural regressions with a framework of two simultaneous linear models, allowing the most basic property of accounting – double entry bookkeeping – to be incorporated as a constraint. The paper aims to investigate whether the constrained seemingly unrelated regression (SUR) estimator with two simultaneous models, produces lower out-of-sample prediction errors than each standalone model. We also examine if CFO and accruals are more capable of predicting future earnings than income statement earnings and expenses. Our findings show that in predicting earnings: (1) a system of structural regressions with two constrained simultaneous models produces significantly smaller out-of-sample prediction errors than each separate regression; and (2) accruals and CFO produce smaller out-of-sample prediction errors than earnings and expenses.

# 1 Introduction

Although research into estimating earnings from other financial statement information continues unabated (e.g., Sloan 1996; Barth et al. 1999; Barth et al. 2001; Dechow et al. 2008; Dechow et al. 2008; Dechow et al. 2010; Lev et al. 2010; Arthur et al. 2010), the application of a suitably constrained model to capture full accounting identities in relation to earnings is relatively new. As documented by Christodoulou and McLeay (2014), empirical research relies heavily on modelling the variation in accounting variables generated by double entry bookkeeping; moreover, accounting variables observed in the same financial statements are contemporaneously codetermined through the resolution of multiple accounting identities. The structural system presented in that study is a generalized framework and can be applied to estimation in any model relying on accounting variables. The structural framework's key feature is the accounting double entry constraint, included to ensure that estimates converge to their theoretically expected relationships. Christodoulou (2018) documents that, based on the double entry constraint, a marginal adjustment to any one accounting input must result in an equal adjustment across all other related variables. In another study, Khansalar and McLeay (2016) explore a model design for explicitly articulated financial statement variables in predicting earnings components. The estimation uses a system of structural regressions, in which the framework of simultaneous linear equations allows the double entry constraint to be incorporated within the model that recognizes the zero-sum articulation of financial statement variables.

Motivated by the above studies' use of a structural system for models reliant on accounting variables, we use a structural system to determine whether and the extent to which imposing a double entry constraint enhances prediction of one-year-ahead earnings out-of-sample. We demonstrate that, based on articulation of financial statements, when double entry is applied consistently to all transactions, and the financial statements are fully articulated, it will be theoretically possible to calculate an item such as earnings not only directly by using income statement, but also by adding CFO and accrulas. Therefore, first, consistent with Sloan (1996), earnings are calculated as CFO plus accruals (EAR_CA model); in the second, earnings are calculated as the difference between expenses and revenues (EAR_ER model). The fact that both equations, by definition, reconcile to the same earnings numbers supports the idea that by using the seemingly unrelated regression (SUR) estimator: (1) regression errors are permitted to be correlated; and (2) estimates converge to their theoretically expected relationships when imposing the accounting double entry bookkeeping constraint.

Christodoulou and McLeay (2014) argue that book value and earnings are endogenously determined (e.g., depreciation expense will be recorded at the current period and subsequent period when a fixed asset is acquired); therefore, they employ double entry constraint to mitigate endogeneity concern inherent in value relevance studies in which price is usually expressed as function of book value and earnings. In the context of earnings prediction model where both sides of the equation are related to earnings, endogeneity does not seem to be a concern. When the EAR_ER and EAR_CA models are simultaneously estimated using SUR, the estimation procedure permits the EAR_ER's revenues and expenses to be estimated contemporaneously with the EAR_CA's CFO and accrual items. Using the SUR estimator, coefficients are estimated simultaneously with the impact of the fundamental variables governing the double entry constraint. The SUR estimator assumes cross-equation errors are correlated, whereas using separate OLS omits useful information obtained through the double entry constraint in articulated financial statements, which can be absorbed by their linked error terms.

We investigate two interrelated research questions that, taken together, aim to extend our understanding of earnings persistence. The first question asks whether using SUR estimation on joint models with an imposed double entry constraint produces lower out-of-sample prediction errors compared to each separate OLS. The SUR estimator produces more efficient estimates than simple OLS when the equations are non-identical and non-nested (Zellner 1962a; Zellner and Huang 1962; Zellner 1963). However, when using the SUR estimator, the effect of imposing the double entry constraint on out-of-sample prediction errors cannot be predicted. This contrasts with in-sample prediction errors: for a given constraint, the errors obtained when the constraint is not imposed are guaranteed to be no larger than when it is.

Barth et al. (2005) document two reasons why imposing a constraint can result in smaller out-of-sample prediction errors. First, other things being equal, using knowledge of the interrelation of accounting amounts in structuring the constraint enhances the models' ability to predict earnings. Second, imposing the constraint mitigates the extent to which the earnings equation overfits the data. However, imposing the constraint can result in larger out-of-sample prediction errors due to its inefficiency in estimating additional forecasting parameters. To test the first research question, we compare the absolute percentage forecast error (APFE) (also known as the mean absolute percentage error (MAPE)), symmetric mean absolute percentage error (SMAPE), adjusted Theil's U-statistic, log of likelihood, Akaike information criterion (AIC), and Bayesian information criterion (BIC) estimations when the constraint is imposed to those when it is not. We find that when the accounting double entry bookkeeping constraint is imposed and both EAR_ER and EAR_CA are

3

simultaneously estimated, out-of-sample prediction errors are significantly smaller than when estimating each model separately.

As the second research question, we investigate whether using SUR estimation with the imposed double entry constraint produces lower out-of-sample prediction errors for EAR_CA in comparison with EAR_ER. Although using CFO and accruals to predict earnings is well documented in the literature, no previous attempt has been made, to our knowledge, to predict earnings using expenses and revenues. Using the above-mentioned tests, we check whether the models' predictive abilities are statistically different. We find that the EAR_CA model produces lower out-of-sample prediction errors than the EAR_ER model, meaning that income statement items are less able to predict future earnings in comparison with CFO and accruals. Khansalar and Namazi (2017) study documents that all income statement items persist more than balance sheet items; they attribute this differential persistence to the greater volume of information contained in income statement variables, which include both accruals and CFO information. Aggregate earnings components in EAR_ER mask this information, and disaggregating earnings information in EAR_ER into accrual and CFO in EAR_CA relaxes the constraint, thus resulting in higher predictive ability. Our results support the position of the FASB/IASB[1] Financial Statement Presentation project regarding the quality of financial statements: it needs to be improved due to the extreme degree of aggregation and netting of items in these statements.

Our study contributes to the extant literature in two ways. First, we document that by using the SUR estimator with two simultaneous equations (EAR_CA and EAR_ER) reconciling to the same earnings numbers, regression errors are permitted to be correlated and, therefore, earnings can be predicted with higher precision. This is because each of the simultaneous equations separately omits useful information obtained through the double entry constraint, which are absorbed by the simultaneous equations' linked error terms. This result confirms the findings of Christodoulou and McLeay (2014) that OLS is unable to cope with highly structured information and is susceptible to simultaneity bias. Importantly, OLS inherently fails to generate estimates adhering to the governing accounting identity.

Second, this research extends the existing literature on earnings persistence since, as noted above, there seems to have been no previous attempt to predict future earnings using income statement expenses and

---

[1] Financial Accounting Standard Board (FASB), International Accounting Standard Board (IASB)

4

revenues. Moreover, we then compare the EAR_ER model's predictions with those of the well-documented EAR_CA model, finding that CFO and accruals better predict future earnings than income statement variables.

This paper is organized as follows. The next section discusses the research design, followed by a section on data and descriptive statistics. The subsequent section reports the empirical results, before the final section then concludes the paper.

## 2 Research Design

As explained by Ijiri et al. (1963), Mann (1984), Fischer et al. (2008), and Christodoulou and McLeay (2014), the articulation of financial statements is an inherent outcome of the accounting double entry bookkeeping system, as every transaction recorded in an account will always be mirrored in one or more other accounts as debit and credit entries. Thus, when double entry bookkeeping is applied consistently to all transactions, and the financial statements are fully articulated, it is theoretically possible to calculate an item such as earnings not only directly from the income statement (EAR_ER model) but also by combining CFO with accruals (EAR_CA model) (Sloan (1996); Barth et al. (2001)).

The following section demonstrates the two models that, by definition, reconcile to the same earnings numbers.

### 2.1 Earnings calculated as CFO plus accruals or revenues minus expenses

In the first model (EAR_CA), we denote a firm's underlying periodic earnings to be calculated as CFO plus accruals. In the context of earnings persistence research, the most basic acknowledgement of the relation between earnings, accruals, and CFO is already implicit in the work of Sloan (1996). This is stated as

$$EAR\_CA_{t+1} = \propto_0 + \beta_1 CFO_t + \beta_2 ACC_t + e_t \qquad (1)$$

Barth et al. (2001) document that each CFO and accrual component reflects different information and that aggregate CFO and accruals mask this information. Disaggregating accruals and CFO into their major components significantly enhances predictive ability.

Following Arthur et al. (2010), CFO can be decomposed into the following components:

5

$$CFO = REC + PAY + TAX + INP + INR + DIV + OTP + OTR, \qquad (2)$$

where REC is cash receipts from customers; PAY is payment to suppliers; TAX is income tax paid; INP is interest paid; INR is interest received; DIV is dividend paid; OTP is all other disclosed cash outflow components not included in the above; and OTR is all other disclosed cash inflow components not included in the above (e.g., excise tax paid).

The first four are included in CFO in the IAS 7 illustrative example; the next two (INR and DIV) may be included in CFO but must be disclosed separately (IAS 7, para 31); and the last two capture the remaining CFO item disclosures.

In addition, following Barth et al. (2001), Baker et al, (2018), Chan et al. (2004),XU et al. (2018) and Eng et al. (2012), accruals (ACC) can be decomposed into the following components:

$$ACC = \Delta ARE + \Delta APA + \Delta INV + DDA + OAC, \qquad (3)$$

where $\Delta ARE$ is change in accounts receivable; $\Delta APA$ is change in accounts payable; $\Delta INV$ is inventory changes; DDA is depreciation, depletion, and amortization; and OAC is other accruals, calculated as Earnings - CFO $-\Delta ARE$ - $\Delta APA$ - $\Delta INV$ – DDA. Barth et al. (2001) documents OAC is the aggregate of other accruals that includes other non-operating activities.

Therefore, equation (1) can be restated as:

$$EAR\_CA_{t+1} = \propto_0 + \beta_1 REC_t + \beta_2 PAY_t + \beta_3 TAX_t + \beta_4 INP_t + \beta_5 INR_t + \beta_6 DIV_t + \beta_7 OTP_t +$$
$$\beta_8 OTR_t + \beta_9 \Delta ARE_t + \beta_{10}\Delta APA_t + \beta_{11}\Delta INV_t + \beta_{12}DDA_t + \beta_{13}OAC_t + e_t \qquad (4)$$

In the second model (EAR_ER), earnings are a company's income after deduction of all day-to-day expenses incurred in the normal course of business, calculated as the difference between revenues and expenses.

$$EAR\_ER_{t+1} = \propto_0 + \beta_1 SAL_t + \beta_2 CGS_t + \beta_3 DDA_t + \beta_4 SAE_t + \beta_5 OOE_t + e_t \qquad (5)$$

where SAL is sales; CGS is cost of goods sold; DDA is depreciation, depletion, and amortization; SAE is selling and administrative expenses; and OOE is other operating expenses.

6

As equations (4) and (5) both reconcile to the same earnings numbers, based on the double entry bookkeeping system, we can impose the following constraint when simultaneously estimating the two models using the SUR estimator:

$$(REC + PAY + TAX + INP + INR + DIV + OTP + OTR) + (\Delta ARE + \Delta APA + \Delta INV + DDA + OAC)$$
$$\equiv (SAL + CGS + DDA + SAE + OOE) \qquad (6)$$

EAR_CA is equal to EAR_ER, this is because in model (3) OAC is the aggregate of other accruals which includes other non-operating activities. Therefore, dependent variables are not different and the fact that the scope of cash flows under EAR-CA is limited to cash flows from operation does not violate the model (6) as OAC acts as a reconciling item. Also to run SUR regression both sides of the model (6) must be equal.

## 2.2 Estimation efficiencies

While researchers often use analysts' forecasts to capture expectations about future performance, concerns about accuracy, bias, and lack of coverage make the use of analysts' forecasts less desirable. This study uses out-of-sample prediction to determine whether the constrained SUR estimator with two simultaneous models, reconciling to the same earnings numbers, enhances predicting one-year-ahead earnings values.

First, we determine to what extent each separate model (EAR_ER and EAR_CA) estimated using the OLS estimator enhances prediction of contemporaneous earnings out-of-sample. Using OLS, all the time series observations are pooled without imposing the constraint and without adjusting for residual correlations. Christodoulou and McLeay (2014) document that regression models estimated using OLS are under-identified and unable to locate unique point estimates for the individual coefficients. They note when one contemporaneous accounting variable is regressed on another, the OLS estimator suffers some simultaneity bias due to the endogeneities arising in the bookkeeping process. OLS is also unable to place the required parameter structure on the endogenous predictor variable, which is required to obtain estimates consistent with the underlying accounting relationship.

Evans et al. (2017) and Hou et al. (2012) provide evidence suggesting that a cross-sectional profitability forecasting model which incorporates the reversion of profitability to expected levels using methods that alleviate the effect of influential observations can lead to forecast accuracy improvement. They extend the literature regarding reliable, accurate, and value-relevant forecasts by employing least absolute deviation (LAD)

analysis instead of ordinary least squares (OLS) because the former approach is able to better accommodate influential observations.

They document that one advantage of the LAD approach, compared with OLS, is that extreme values are less influential; specifically, LAD minimizes the sum of absolute errors, rather than squared errors, as in OLS. Accordingly, their estimation method could incorporate small firms and firms with frequent or large losses.

Thus, they develop a feasible method of generating reliable ex ante earnings forecasts for these firms that differs from currently adopted approaches in this stream of research. Results reveal that forecasts from their model are more accurate than extant models at every forecast horizon considered.

They find that their model is also significantly more accurate than a scaled version of the cross-sectional model proposed by Hou et al. (2012) for forecast horizons of one to five years.

In the current study, influential observations are not eliminated. We only exclude firm-year observations with nonsensical signs, for instance, companies with original negative sales, firm-year observations that did not pass the double entry verification check due to input errors during data processing and outliers.

In this study, when the EAR_CA and EAR_ER models are estimated using SUR, the EAR_ER's revenues and expenses can be estimated contemporaneously with the EAR_CA's CFO and accrual items; moreover, the coefficients can be estimated simultaneously with the impact of the fundamental variables governing the double entry constraint.

The SUR estimator assumes that, the sole link between models being channeled through their error terms, and if the ordinary Least Square (OLS) estimator is used, useful information obtained by the double entry constraint omits which can be absorbed by their linked error terms. We ran the Breusch and Pagan (1980) and the Harvey and Phillips (1982) tests to check if there is contemporaneous dependence between the cross-equation error terms.

The SUR system also allows the imposition of cross-equation parameter constraints, as required by the analytical framework. As stated above, when using the SUR estimator, the effect of imposing the double entry constraint on out-of-sample prediction errors cannot be predicted, as although the interrelation of accounting amounts enhances the predicting ability, imposing the constraint overfits the data. Therefore, due to inefficiency

8

in estimating the additional forecasting parameters, imposing the constraint can result in larger out-of-sample prediction errors when models are estimated using SUR.

One might also expect earnings prediction errors to decrease as the level of earnings disaggregation increases: for example, in equation (1), which disaggregates accruals and CFO into their underlying components. This is because, as the level of earnings disaggregation increases, different earnings components are permitted to have different valuation multiples. However; earnings disaggregation can be costly for SUR in terms of increasing prediction errors. Initially, out-of-sample prediction errors can increase as the level of earnings disaggregation increases, due to the potential for data overfitting. However, as the level of earnings disaggregation increases, so too does the extent of structure imposed by the constraint on the forecasting earnings. In other words, although earnings disaggregation relaxes constraints on valuation coefficients by permitting them to differ, it adds constraints on the earnings components coefficients when the accounting double entry bookkeeping constraint is imposed.

In the out-of-sample prediction process, we obtain earning predictions for each firm without using that firm's data to generate its predicted earnings. The prediction of firm i's earnings in year t+1 is the value predicted from EAR_CA or/and EAR_ER using estimated coefficients and all firms' data except firm i's in year t+1. Because firm i's data in year t+1 are not used to estimate the coefficients, each prediction is out-of-sample.

Using the OLS estimator, when generating the earning prediction for firm i in year t+1 without imposing the double entry constraint, we estimate either EAR_CA or EAR_ER using the data for all firms except firm i in year t+1. However, using the SUR estimator when generating the earnings prediction for firm i in year t+1, we estimate EAR_CA and EAR_ER using the data for all firms except firm i in year t+1 and restricting the coefficients of EAR_CA to equal those implied by EAR_ER, in the manner specified by the accounting double entry bookkeeping constraint.

**2.3  Prediction Error Tests**

Consistent with the studies of Francis and Eason (2012), Krishnan and Largay III (2000), Burgstahler and Dichev (1997), Kim and Kross (2005), Lorek and Willinger (2009), and Lev et al. (2010), to compare the forecast accuracy of alternative models, we use a variety of measures such as the APFE, SMAPE, adjusted

Theil's U-statistic, log of likelihood, AIC, and BIC estimations to test whether our results are sensitive to the method used. The initial evaluation of forecast accuracy, APFE, is calculated as follows:

$$\text{APEF}_{\text{OLS/SUR}} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{\text{Actual EAR\_CA/ER}_{t+1} - \text{Forecast EAR\_CA/ER}_{t+1}}{\text{Actual EAR\_CA/ER}_{t+1}} \right|$$

The absolute value in this calculation is summed for every forecasted point in time and divided by the number of fitted points n, multiplying by 100 to make it a percentage error.

The study's first question asks whether using the constrained SUR produces lower out-of-sample prediction errors than using each OLS regression separately. Therefore, each observation in the sample generates four APFEs: two for EAR_CA (OLS and SUR) and two for EAR_ER (OLS and SUR). For example, for EAR_CA, one forecast value arises from each one of the two OLS or SUR estimators and smaller forecast errors indicate greater forecast accuracy. We expect that when the double entry constraint is imposed and both EAR_CA and EAR_ER are simultaneously estimated using SUR, out-of-sample prediction errors are significantly smaller than when separately estimating each model using OLS. Thus, the paired differences between EAR_CA (using OLS) & EAR_CA (using SUR) and EAR_ER (using OLS) & EAR_ER (using SUR) will be positive, on average (i.e., EAR_CA/ER (using OLS) error - EAR_CA/ER (using SUR) error > 0).

We also use SMAPE. It is usually defined as follows:

$$\text{SMAPE}_{\text{OLS/SUR}} = \frac{1}{n} \sum_{t=1}^{n} \frac{\left| \text{Forecast EAR\_CA/ER}_{t+1} - \text{Actual EAR\_CA/ER}_{t+1} \right|}{\left( \left| \text{Actual EAR\_CA/ER}_{t+1} \right| \right) + \left( \left| \text{Forecast EAR\_CA/ER}_{t+1} \right| \right) / 2}$$

The value of this calculation is summed for every fitted point and divided by the number of fitted points n.

We also use the adjusted Theil's U-statistic to measure forecast accuracy. Theil (1966) defines the U-statistic as $\frac{\sqrt{\sum P_t - A_t)^2}}{\sqrt{\sum A_t)^2}}$, where P and A represent changes in predicted and actual values, respectively, which could evaluate the forecast accuracy using a levels-type measurement. The key idea is that Theil's (1966) U-statistic will equal zero if the forecast is perfect (i.e., actual value equals forecast value), and equal unity if not. It is easy to see how researchers possibly misinterpret P and A as levels-type measurements and, therefore, overstate forecast model accuracy, as the use of levels-type measurements artificially decreases the U-statistic's value. Furthermore, levels-type measurements fail to create the (0,1) boundaries intended by Theil (1966).

10

Francis and Eason (2012) introduced an alternative forecast metric that uses Theil's (1966) U-statistic with explicit indicators for change variables: $\frac{\sqrt{\sum(\Delta P_t - \Delta A_t)^2}}{\sqrt{\sum \Delta \Delta A_t)^2}}$, where $\Delta P$ and $\Delta A$ represent changes in the predicted and actual values. As mentioned previously, a no-change forecast (i.e., $\Delta P = 0$) will generate a U-statistic equal to one, while a perfect forecast (i.e., $\Delta P = \Delta A$) will generate a U-statistic equal to zero. Contrasting OLS-estimated EAR_CA/ER with SUR-estimated EAR_CA/ER provides evidence on whether the constrained SUR produces lower out-of-sample prediction errors than does each separate OLS regression.

Finally, we compare log of likelihood, AIC, and BIC between the two OLS and SUR estimators. AIC is a measure of the relative goodness of fit of a statistical model; it describes the trade-off between bias and variance in model construction, or, loosely speaking, between accuracy and complexity in a model. Several candidate models can be ranked according to their AIC values, the best model being that with the lowest AIC. BIC is another criterion for model selection among a finite set of models. When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in overfitting. The BIC resolves this problem by introducing a penalty term for the number of parameters in the model. Therefore, as with the AIC, the model with the lowest BIC is the best.

The study's second question investigates whether using constrained SUR produces lower out-of-sample prediction errors for EAR_CA in comparison with EAR_ER. We run the APFE, SMAPE, and adjusted Theil's U-statistic to test whether, using SUR, our results are sensitive to the method used (EAR_CA or EAR_ER). For example, we check the paired differences between SUR-estimated EAR_CA and SUR-estimated EAR_ER. Finally, the current study re-examines Sloan (1996) study by further investigating whether, in the EAR_CA model, CFO components of earnings have a stronger relationship with future earnings than all accrual components. To test this, we use F-tests of the coefficient equality across the accrual and CFO components. We document that accrual components evidence differential persistence into future earnings.

## 2.4  The double entry constraint

The likely magnitude of the efficiency gain through SUR has been investigated by Zellner (1962a, Zellner (1962b), Revankar (1974), Binkley (1982), and Kmenta (1997); they note that under conditions generally encountered in practice, the regression coefficients obtained using SUR are more efficient than those obtained through equation-by-equation application of least squares.

11

Following Zellner (1962a), we have:

$$y_\mu = X_\mu \beta_\mu + \varepsilon_\mu \tag{7}$$

As demonstrated, $y_\mu$ can be contemporaneously codetermined through the resolution of multiple identities using the equations m = 1,2,…,M. The system may be written as:

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & ... & 0 \\ 0 & X_2 & ... & 0 \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ 0 & 0 & ... & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_M \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_M \end{bmatrix} \tag{8}$$

The $M \times 1$ disturbance vector is assumed to have the following variance-covariance matrix:

$$\Sigma = V(u) = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I & . & . & . & \sigma_{1M}I \\ \sigma_{21}I & \sigma_{21}I & . & . & . & \sigma_{2M}I \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma_{M1}I & \sigma_{M2}I & . & . & . & \sigma_{MM}I \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & . & . & . & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & . & . & . & \sigma_{2M} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma_{M1} & \sigma_{M2} & . & . & . & \sigma_{MM} \end{bmatrix} \otimes I \tag{9}$$

where I is a unit matrix of order $T \times T$, $\sigma_{\mu\mu'} = E(u\_\mu t\ u\_\mu' t)$ for t =1, 2,…,T, and μ and μ' = 1, 2,…, M.

In OLS, $\mu \neq \mu'$ implies that the $\sigma\_\mu\mu'$ are the same for all variables and that there is no correlation between different independent variables' disturbances. However, in the SUR approach it is assumed that all the explanatory variables for the regressions are comprised only of predetermined covariates, the sole link between the equations being channeled through their error terms.

Equation (9) can be restated for the EAR_CA and EAR_ER models in the current study as follows:

$$
\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}
\begin{bmatrix}EAR\_CA_{it+1}\\EAR\_ER_{it+1}\\0\\0\end{bmatrix}
=
\begin{bmatrix}
\beta_1 & \beta_2 & \beta_3 & \cdots & \beta_{11} & \beta_{12} & \beta_{13} & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & \beta_{14} & \beta_{15} & \beta_{16} & \beta_{17} & \beta_{18}\\
\beta_1 & \beta_2 & \beta_3 & \cdots & \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} & \beta_{15} & \beta_{16} & \beta_{17} & \beta_{18}\\
0 & 0 & 0 & \cdots & 0 & \beta_{12} & 0 & 0 & 0 & \beta_{16} & 0 & 0
\end{bmatrix}
\begin{bmatrix}
REC_{it}\\ PAY_{it}\\ TAX_{it}\\ INP_{it}\\ INR_{it}\\ DIV_{it}\\ OTP_{it}\\ OTR_{it}\\ \Delta ARE_{it}\\ \Delta APA_{it}\\ \Delta INV_{it}\\ DDA_{it}\\ OAC_{it}\\ SAL_{it}\\ CGS_{it}\\ DDA_{it}\\ SAE_{it}\\ OOE_{it}
\end{bmatrix}
+
\begin{bmatrix}u_{1it}\\u_{2it}\\0\end{bmatrix}
\quad (10)
$$

Eighteen variables are presented for both EAR_CA and EAR_ER: for instance, in the first row, for **$EAR\_CA_{it+1}$**, we have thirteen coefficients $(\beta_1, \beta_2, \beta_3, ..., \beta_{13})$ for the variables from **$REC_{it}$** to **$AC_{it}$**; in the second row, for **$EAR\_ER_{it+1}$**, we have five coefficients $(\beta_{14}, \beta_{15}, \beta_{16}, \beta_{17}, \beta_{18})$ for the variables from **$SAL_{it}$** to **$OOE_{it}$**.

The third row, zero, identifies the accounting double entry bookkeeping constraint given by the articulation of financial statements. This allows for the comprehensive identification of earnings that imposes the constraint on all eighteen coefficients $(\beta_1, \beta_2, \beta_3, ..., \beta_{18})$ for both EAR_CA and EAR_ER, and assumes that the summation of all thirteen coefficients for EAR_CA is equal to the summation of all five coefficients for EAR_ER. We suppress the output of the estimated intercepts and focus the analysis on the variables' structural coefficients. DDA is employed as an additional explanatory variable for both models. For the EAR_CA, it appears as a component of accruals; for EAR_ER, as an expense. Therefore, we constrain the coefficient for depreciation across the two models to be equal. The last row, zero, shows the constraint on DDA.

## 3    Data and descriptive statistics

The initial sample of US firms was drawn from the Worldscope database. Worldscope contains complete coverage of US companies filing with the Securities Exchange commission, with the exception of closed end investment companies. Also it contains extinct or inactive companies, i.e. those which have merged, liquidated or become privately held. History for these companies remains on the database. Table 1, Panel A documents that there were initially 6,428 firms or 38,285 firm-year observations over the sample period of fiscal years 2004

13

through 2014. Following Melendrez et al. (2008), we eliminate the firms in financial or utility industries because the demarcation between their operating and financing activities is not clear. To minimize the effect of outliers, we follow Bollen and Jackma (1985), Kim and Kross (2005), and Wu and Xu (2008) by using the studentized residual technique to detect outliers. The outliers are presented for each regression separately. By using this technique, an outlier is defined as an observation whose dependent-variable value is unusual given its values on the predictor variables. In other words, studentised residuals are helpful in identifying outliers that do not appear to be consistent with the rest of the data. In the case of studentised residuals, large deviations from the regression line are identified. Since the residuals from a regression will generally not be independently or identically distributed (even if the disturbances in the regression model are), it is advisable to weight the residuals by their standard deviations (this is what is meant by studentisation). In summary, a studentised residual is the quotient resulting from the division of a residual by an estimate of its standard deviation. In this study, any observations whose absolute studentised residuals are two or more in regressions are identified as outliers.

**Table 1 - Panel A**    Data selection

| Construction of the dataset | Number of Firms | Number of observations |
|---|---|---|
| Initial sample identified from Worldscope for fiscal years 2004 to 2014 | 6,428 | 38,285 |
| | | |
| Less: Financial and utility industries | 850 | 4,860 |
|     Outliers for EAR_CA model | 186 | 847 |
|     Outliers for EAR_ER model | 136 | 650 |
|     Observations with nonsensical signs | 59 | 350 |
|     Observations where income statement is not balanced | 38 | 190 |
|     Observations where balance sheet is not balanced | 52 | 296 |
|     Observations where cash flow is not balanced | 24 | 132 |
| Final Sample | 5,083 | 30,960 |

**Panel B**    Descriptive statistics

EAR_CA

| CFO | Mean | Min. | Max. | Std.Dev. |
|---|---|---|---|---|
| REC | 0.2274 | 0.0000 | 0.3315 | 0.0420 |
| PAY | -0.2052 | -0.3294 | -0.0013 | 0.0344 |
| TAX | -0.0064 | -0.0104 | 0.0000 | 0.0311 |
| INP | -0.0021 | -0.0034 | 0.0000 | 0.0258 |
| INR | -0.0013 | -0.0021 | 0.0000 | 0.2981 |
| DIV | -0.0009 | -0.0015 | 0.0000 | 0.0352 |
| OTP | -0.0016 | -0.0026 | 0.0000 | 0.2564 |
| OTR | 0.0014 | 0.0000 | 0.0021 | 0.2159 |
| | 0.0113 | | | |

| Accruals | Mean | Min. | Max. | Std.Dev. |
|---|---|---|---|---|
| $\Delta$ARE | 0.0008 | -0.2652 | 0.3894 | 0.0284 |
| $\Delta$APA | -0.0027 | -0.2406 | 0.2894 | 0.0258 |
| $\Delta$INV | 0.0024 | -0.3521 | 0.3821 | 0.0322 |
| DDA | -0.0048 | -0.2415 | 0.0006 | 0.0289 |
| OAC | -0.0011 | -0.2913 | 0.3298 | 0.0311 |
| | -0.0054 | | | |
| | | 0.0059 | | |

EAR_ER

| Earnings | Mean | Min. | Max. | Std.Dev. |
|---|---|---|---|---|
| SAL | 0.2256 | 0.0000 | 0.3615 | 0.0368 |
| CGS | -0.2053 | -0.3451 | -0.0112 | 0.0415 |
| DDA | -0.0048 | -0.2415 | 0.0006 | 0.0289 |
| SAE | -0.0082 | -0.2135 | -0.0152 | 0.0358 |
| OOE | -0.0014 | -0.2585 | 0.2435 | 0.0325 |
| | | 0.0059 | | |

All variables are deflated by the summation of the absolute value of all 18 variables defining both EAR_CA  and EAR_ER. Double entry bookkeeping applies on the sample means. Min. and Max. indicate the range of variation and Std.Dev. is the standard deviation. See Eq (2), (3) and (5) for variable definitions.


We eliminate firm-year observations with nonsensical signs. That means any observation with positive Cost of Goods Sold or negative Sales Revenue has been dropped from the sample. For example, with regard to Cost of Goods Sold, purchasing raw materials has to be debited into Cost of Goods Sold and credited into any

interrelated account; however, on some occasions, a company may face some unusual circumstances leading to not only stopping buying raw materials for an entire specific year, but also starting to return some of those raw materials to the suppliers. In this unusual situation we decided to drop any observations like this from the sample. Also, companies with original negative sales were dropped from the sample as this signals that a company that has no sales for the current year but a sales return from the previous year.

Every transaction recorded in an account will always be mirrored in one or more other accounts as debit and credit entries. Thus, following Christodoulou and McLeay (2014), when the accounting double entry bookkeeping constraint is applied consistently to all transactions and financial statements, it is theoretically possible to calculate an item in several ways.

Before we use the accounting double entry condition to constrain simultaneous models, we run a double entry verification check to ensure that the accounts add up and that there was no input error during data processing (either from the preparer or the database provider). It is also known that various measurement errors may arise due to the computation of accruals from individual transactions under incomplete double entry (Hribar and Collins (2002). We ran this verification check for balance sheets, income statements, and cash flow statements. The final sample comprises 5,083 firms, producing 30,960 firm-year observations.

Table 1, Panel B gives the standard statistical summary for the variables used in the estimation. We scale all variables by the summation of the absolute value of all 18 variables defining both EAR_CA and EAR_ER. The applied deflator in this study has the lowest level of multicollinearity compared to other possible deflators, e.g., invested assets.[2]

**In Panel B, the double entry constraint holds in the arithmetic means of the earnings identity, EAR_CA ≡ EAR_ER (i.e., 0.0113 + -0.0054 ≡ 0.0059). Consistent with prior research, e.g., Sloan (1996), the means of earnings and CFO are positive and those of aggregate accruals (accruals= earnings – CFO) are negative.** Pearson and Spearman correlation coefficients were calculated for all regression variables, but, for brevity, the tables are not presented.[3]

---

[2] Following Fairfield et al. (2003), we do not use invested assets as a deflator. Their findings suggest that the lower persistence of accruals, compared to cash flows, is due to accruals being highly correlated with the growth of invested assets, as used in prior research as the denominator to scale future earnings.

[3] As our matrix has 18 variables, for brevity, we only report the outcomes. Correlation coefficients between REC and PAY as well as SAL and CGS were particularly high, which could signal potential multicollinearity problems. To test for multicollinearity, we ran a diagnostic test: variance inflation factor (VIF). Chatterjee and Hadi (2015) and Baum et al. (2003) recommend a maximum VIF of 10, above which the estimates are too sensitive to even small changes in the data (i.e., unstable). We found the highest VIFs to be 3.21 for SAL and 3.01 for CGS, which are well below the aforementioned maximum recommended VIF. Binkley (1982) notes that some degree of multicollinearity is unavoidable, especially in accounting models that rely on such highly structured information.

## 4     Empirical Results

As stated above, the fact both EAR_CA and EAR_ER, by definition, reconcile to the same earnings numbers lends support to the idea that regression errors are correlated and all explanatory variables comprise pre-determined regressors; moreover, the link between the equations is channeled through their error terms whereas using each EAR_CA or EAR_ER separately omits useful information obtained by articulation of financial statements. Consistent with Christodoulou and McLeay (2014), we test whether regression errors are correlated by using Breusch and Pagan (1980) Lagrange Multiplier test, which examines asymptotically the null hypothesis of no cross-equation error correlation $E\left(U_{1\,it}\,U_{2\,it}\right)$ in equation (10). In addition, we run Harvey and Phillips (1982) exact finite independence test between one regression and the residual term of the other regression. Both tests verify the strong contemporaneous dependence between the cross-equation error terms at $p < 0.0001$. This validates use of the SUR system for estimating the joint regressions of earnings.

Table 2 presents regression results for both EAR_CA and EAR_ER using OLS and the SUR estimator with the accounting double entry equilibrium condition. The OLS model is conceptually under-identified because it lacks the predetermined equilibrium condition. Indeed, including the accounting double entry constraint guides estimation toward equilibrium (with the sum of the estimated expected values equaling the sum of the estimated coefficients), and reveals where the OLS struggles to find a more precise solution. It has been argued throughout that the accounting double entry constraint underlines a deterministic relationship between the 18 variables across the two models, as connected by the equilibrium condition of EAR_CA is equal to EAR_ER. Therefore, theory predicts that the addition of all the estimated expected values of EAR_CA will be identical to those of EAR_ER. We define estimated expected values as the multiplication of the estimated coefficients reported in Table 2 by the mean expected values reported in Table 1, Panel B, which adds up to 0.0473 [8.02 × 0.0059] for both models. Using OLS, the estimated expected values for the two models, which should be identical, are not so. More importantly, as OLS cannot ensure that estimates follow the accounting double entry rules, it should not qualify for interpretation, just as an unbalanced balance sheet would not qualify as a reliable statement of financial position.

17

**Table 2**  Regression results under OLS and SUR

$$EAR_{CA_{t+1}} = \alpha_0 + \beta_1 REC_t + \beta_2 PAY_t + \beta_3 TAX_t + \beta_4 INP_t + \beta_5 INR_t + \beta_6 DIV_t + \beta_7 OTP_t + \beta_8 OTR_t +$$
$$\beta_9 \Delta ARE_t + \beta_{10} \Delta APA_t + \beta_{11} \Delta INV_t + \beta_{12} DDA_t + \beta_{13} OAC_t + e_t \tag{5}$$

$$EAR\_ER_{t+1} = \alpha_0 + \beta_1 SAL_t + \beta_2 CGS_t + \beta_3 DDA_t + \beta_4 SAE_t + \beta_5 OOE_t + e_t \tag{7}$$

| Vars | OLS _ EAR_CA | | OLS _ EAR_ER | | SUR | |
|---|---|---|---|---|---|---|
| | Coeff | Std. Err. | Coeff | Std. Err. | Coeff | Std. Err. |
| **CFO** | | | | | | |
| REC | 0.42 | 0.0188 | | | 1.88*** | 0.013 |
| PAY | 0.56*** | 0.0181 | | | 0.61*** | 0.016 |
| TAX | 0.54*** | 0.0225 | | | 0.66*** | 0.013 |
| INP | 0.55*** | 0.0182 | | | 0.66*** | 0.014 |
| INR | 0.64* | 0.0158 | | | 0.63*** | 0.012 |
| DIV | 0.54*** | 0.0135 | | | 0.56*** | 0.011 |
| OTP | 0.83*** | 0.0132 | | | 0.44*** | 0.011 |
| OTR | 0.44* | 0.0185 | | | 0.48*** | 0.013 |
| **Accruals** | | | | | | |
| ΔARE | 0.35 | 0.0125 | | | 0.45*** | 0.0096 |
| ΔAPA | 0.36*** | 0.0104 | | | 0.87*** | 0.0029 |
| ΔINV | 0.34*** | 0.0182 | | | 0.89*** | 0.0024 |
| DDA | 0.80 | 0.0084 | | | 0.51*** | 0.0084 |
| OAC | 0.34*** | 0.0075 | | | 0.65*** | 0.0055 |
| | 6.71 | | | | 8.02 | |
| **Earnings** | | | | | | |
| SAL | | | 0.76** | 0.0037 | 1.98*** | 0.0023 |
| CGS | | | 0.43*** | 0.0094 | 1.88*** | 0.0046 |
| DDA | | | 0.35** | 0.0084 | 0.51*** | 0.0084 |
| SAE | | | 0.67*** | 0.005 | 0.51*** | 0.0016 |
| OOE | | | 0.59* | 0.0075 | 1.87*** | 0.0061 |
| | | | 2.19 | | 8.02 | |
| Adj. $R^2$(%) | 11 | | 13 | | 34 | |

\*, \*\*, and \*\*\* indicate a significant difference from zero on a ten, five, and one percent level, respectively.

Consistent with Sloan (1996), Barth et al. (1999), Pfeiffer and Elgers (1999), Collins and Hribar (2000), and Melendrez et al. (2008), we find that, when using OLS, that accrual components of earnings are less persistent than CFO components. With 99% confidence attributed to all coefficients, accrual components are less persistent than CFO components of earnings. To test this, we used F-tests of coefficient equality across the accrual components and CFO components, and, with 99% confidence, reject the hypothesis that the coefficients of accrual components are equal to those of CFO components.

To test the study's first research question, Table 3, Panel A, presents out-of-sample prediction results using APFE, SMAPE, adjusted Theil's U-statistic, log of likelihood, AIC, and BIC estimations when the constraint is imposed compared to those when it is not. All the indicators show that out-of-sample prediction errors are significantly smaller when using constrained SUR, rather than estimating each model separately.

18

**Table 3- Panel A**    Out-of-sample prediction results (EAR_CA)

| Median APFE | | Median paired forecast error difference | SMAPE | | Symmetric mean paired error difference | Theil's U-statistic |
|---|---|---|---|---|---|---|
| OLS | SUR | OLS & SUR | OLS | SUR | OLS & SUR | OLS & SUR |
| 74.6% | 67.8% | **6.3%**[***] | 42.3% | 38.7% | **3.6%**[***] | **0.834** |

Out-of-sample prediction results (EAR_ER)

| Median APFE | | Median paired forecast error difference | SMAPE | | Symmetric mean paired error difference | Theil's U-statistic |
|---|---|---|---|---|---|---|
| OLS | SUR | OLS & SUR | OLS | SUR | OLS & SUR | OLS & SUR |
| 65.6% | 58.8% | **7.7%**[***] | 39.3% | 35.5% | **4.1%**[***] | **0.916** |

| | OLS | | SUR |
|---|---|---|---|
| | EAR_CA | EAR_ER | |
| Log-likelihood | 38,452 | 54,233 | 140,325 |
| AIC – Akaike information criterion | -76,653 | -98,258 | -325,215 |
| BIC – Bayesian information criterion | -78,254 | -68,258 | -315,659 |

n = 30,960

*, **, and *** indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

**Panel B**    Out-of-sample SUR prediction results (EAR_CA) and (EAR_ER)

| Median APFE | | Median paired forecast error difference | SMAPE | | Symmetric mean paired error difference | Theil's U-statistic |
|---|---|---|---|---|---|---|
| EAR_ER | EAR_CA | EAR_ER vs EAR_CA | EAR_ER | EAR_CA | EAR_ER & EAR_CA | EAR_ER & EAR_CA |
| 69.1% | 65.3% | **4.3%**[***] | 44.3% | 36.5% | **6.4%**[***] | **0.815** |

n = 30,960

*, **, and *** indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

The APFE, SMAPE, and adjusted Theil's U-statistic confirm that when constrained SUR is used, the EAR_CA model's forecast measures are significantly more accurate than EAR_CA's measures using OLS. For example, the SUR-estimated EAR_CA's median APFE is, on average, 6.3% more accurate than that of the OLS-estimated EAR_CA. Furthermore, the SUR-estimated EAR_CA's SMAPE is, on average, 3.6% more accurate than that of the OLS-estimated EAR_CA. Note that the corresponding Theil's U-statistic for the SUR-estimated EAR_CA over the OLS-estimated EAR_CA is 0.834. Comparison of the log of likelihood, AIC, and BIC between the two models shows that constrained SUR is more efficient than the OLS estimators. Despite the SUR system consuming more degrees of freedom than the single-equation OLS, the gain in terms of likelihood is much greater. In our analysis, the OLS estimators are not as efficient as the SUR model. As with the AIC, the model with the smallest BIC, here SUR, is more efficient than OLS.

Testing the study's second research question, we find that the EAR_CA model produces lower out-of-sample prediction errors than the EAR_ER model, meaning that CFO and accruals are better able to predict future earnings than income statement items. To test this, Table 3, Panel B presents the APFE, SMAPE, and adjusted Theil's U-statistics to compare predictive ability between EAR_CA and EAR_ER. All of the indicators confirm that EAR_CA prediction errors are significantly smaller than those of EAR_ER. For example, when estimating both models jointly under constrained SUR, EAR_CA's median APFE is, on average, 4.3% more accurate than that of EAR_ER. Furthermore, EAR_CA's SMAPE is, on average, 6.4% more accurate than that of EAR_ER. The corresponding Theil's U-statistic for EAR_CA over EAR_ER is 0.815.

## 5    Conclusion

In this study, we investigated three interrelated research questions concerning the double entry constraint in joint regressions, aiming to extend our understanding of its usefulness in predicting earnings. We were motivated to pursue this study by Christodoulou and McLeay (2014), Christodoulou (2017), and Khansalar and McLeay (2016), as the structural system presented in their studies is a generalized framework and can be applied to estimation in any model relying on accounting variables. We use a structural system to determine whether and the extent to which imposing the double entry constraint enhances predicting one-year-ahead earnings out-of-sample.

Using SUR estimation, we find that constrained joint regressions produce lower out-of-sample prediction errors compared to each separate OLS estimator. This is because both applied models, by definition, reconcile to the same earnings numbers and, therefore, regression errors are permitted to be correlated; consequently, estimates converge to their theoretically expected relationships according to the accounting double entry bookkeeping constraint. In fact, it is assumed that, using SUR, the sole link between the regressions is channeled through their error terms; in contrast, each of the separate OLS regressions omits useful information obtained by the double entry constraint.

Our results also suggest that, using constrained SUR estimation, CFO and accruals are better able to predict future earnings compared to income statement revenues and expenses. It is interesting that while financial performance is customarily assessed by summarizing a business's revenues and expenses, we find that CFO and accruals produce lower out-of-sample prediction errors regarding future earnings. One potential explanation is the greater volume of information yielded by disaggregating revenues and expenses into accruals and CFO. This

20

finding supports the position of the FASB/IASB Financial Statement Presentation project on the quality of financial statements, namely that it should be improved due to the extreme degree of aggregation and netting of items in these statements.

Taken together, our results offer evidence that using the double entry constraint mitigates out-of-sample prediction errors and enhances the accuracy of earnings forecasts.

## References

Arthur, N., Cheng, M. & Czernkowski, R. (2010). Cash flow disaggregation and the prediction of future earnings. Accounting & Finance 50:1-30. https://doi.org/10.1111/j.1467-629x.2009.00316.x

Baker, T.A., Lopez, T.J., Reitenga, A.L. and Ruch, G.W., (2018). The influence of CEO and CFO power on accruals and real earnings management. Review of Quantitative Finance and Accounting 1-21. https://doi.org/10.1007/s11156-018-0711-z

Barth, M. E., Beaver, W. H., Hand, J. R. & Landsman, W. R. 1999. Accruals, cash flows, and equity values. Review of Accounting Studies, 4, 205-229. https://doi.org/10.2139/ssrn.149768

Barth, M. E., Beaver, W. H., Hand, J. R. & Landsman, W. R. 2005. Accruals, accounting-based valuation models, and the prediction of equity values. Journal of Accounting, Auditing & Finance, 20, 311-345. https://doi.org/10.1177/0148558x0502000401

Barth, M. E., Cram, D. P. & Nelson, K. K. 2001. Accruals and the prediction of future cash flows. The Accounting Review, 76, 27-58. https://doi.org/10.1108/01409171111146715

Baum, C. F., Schaffer, M. E. & Stillman, S. 2003. Instrumental variables and GMM: Estimation and testing. Stata journal, 3, 1-31.

Binkley, J. K. 1982. The effect of variable correlation on the efficiency of seemingly unrelated regression in a two-equation model. Journal of the American Statistical Association, 77, 890-895. https://doi.org/10.2307/2287323

Bollen, K. A. & Jackma, R. N. 1985. Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases. Sociological Methods and Research, 13, 510-42.

Breusch, T. S. & Pagan, A. R. 1980. The Lagrange multiplier test and its applications to model specification in econometrics. The Review of Economic Studies, 47, 239-253. https://doi.org/10.2307/2297111

Burgstahler D, Dichev I, (1997) Earnings management to avoid earnings decreases and losses. Journal of accounting and economics 24: 99-126. https://doi.org/10.1016/s0165-4101(97)00017-7

Carlson R L, (1978) Seemingly unrelated regression and the demand for automobiles of different sizes, 1965-75: A Disaggregate Approach. The Journal of Business 51: 243-262. https://doi.org/10.1086/295996

Chan K, Jegadeesh N, Sougiannis T, (2004) The accrual effect on future earnings. Review of Quantitative Finance and Accounting 22:97-121. https://doi.org/10.1023/b:requ.0000015852.00973.8f

Chatterjee S, Hadi A S, 2015 Regression analysis by example. John Wiley & Sons.

Christodoulou D, (2018) The accounting identity trap: identification under stock-and-flow rank deficiency. Applied Economics, 50:1413-1427. https://doi.org/10.1080/00036846.2017.1363860

Christodoulou D, Mcleay S, (2014) The Double Entry Constraint, Structural Modeling and Econometric Estimation. Contemporary Accounting Research 31:609-628. https://doi.org/10.1111/1911-3846.12038

Collins D W, Hribar P, (2000) Earnings-based and accrual-based market anomalies: one effect or two? Journal of Accounting and Economics 29:101-123. https://doi.org/10.2139/ssrn.166455

Dechow P, Ge W, Schrand C, (2010) Understanding earnings quality: A review of the proxies, their determinants and their consequences. Journal of accounting and economics 50:344-401. https://doi.org/10.1016/j.jacceco.2010.09.001

Dechow P M, Richardson S A, Sloan R G, (2008) The persistence and pricing of the cash component of earnings. Journal of Accounting Research 46:537-566. https://doi.org/10.2139/ssrn.638622

Eng L L, Lin Y C, (2012) Accounting quality, earnings management and cross-listings: evidence from China. Review of Pacific Basin Financial Markets and Policies 15:1250009. https://doi.org/10.1142/s0219091512500099

Evans M E, Njoroge K, Yong, K.O.. (2017) An examination of the statistical significance and economic relevance of profitability and earnings forecasts from models and analysts. Contemporary Accounting Research 34:1453-1488 https://doi.org/10.1111/1911-3846.12307

Fairfield P M, Whisenant J S, Yohn T L, (2003) Accrued earnings and growth: Implications for future profitability and market mispricing. The accounting review 78:353-371. https://doi.org/10.2308/accr.2003.78.1.353

Fischer M, Gordon T P, Khumawala S B, (2008) Tax-Exempt Organizations and Nonarticulation: Estimates Are No Substitute for Disclosure of Cash Provided by Operations. Accounting Horizons 22:133-158. https://doi.org/10.2308/acch.2008.22.2.133

Francis R N, Eason P, (2012) Accruals and the naïve out-of-sample prediction of operating cash flow. Advances in Accounting 28:226-234. https://doi.org/10.1016/j.adiac.2012.09.004

Giles D E, (2004) Calculating a standard error for the Gini coefficient: some further results. Oxford Bulletin of Economics and Statistics 66:425-433. https://doi.org/10.1111/j.1468-0084.2004.00086.x

Harvey A, Phillips G, (1982) The estimation of regression models with time-varying parameters. Games, economic dynamics, and time series analysis. Springer 306-321. https://doi.org/10.1007/978-3-662-41533-7_18

Hou K, Van Dijk M A, Zhang, Y.. (2012) The implied cost of capital: A new approach. Journal of Accounting and Economic, 53:504-526. https://doi.org/10.2139/ssrn.1343516

Hribar P, Collins D W, (2002) Errors in estimating accruals: Implications for empirical research. Journal of Accounting research 40:105-134. https://doi.org/10.1111/1475-679x.00041

Ijiri Y, Levy F, Lyon R, (1963) A linear programming model for budgeting and financial planning. Journal of Accounting Research 198-212. https://doi.org/10.2307/2489855

Khansalar E, Mcleay S, (2016) Articulated Accounting: Accruals, Cash Flow and the Double Entry Constraint on Earnings. Accounting, Finance and Governance Review 23:19-41

Khansalar E, Namazi M, (2017) Cash flow disaggregation and prediction of cash flow. Journal of Applied Accounting Research 18: 464-479 https://doi.org/10.1108/jaar-02-2015-0011

Kim M, Kross W,( 2005) The ability of earnings to predict future operating cash flows has been increasing—not decreasing. Journal of Accounting Research 43:753-780. https://doi.org/10.2139/ssrn.303283

Kmenta J, (1997) The elements of econometrics. Macmillan Co. https://doi.org/10.3998/mpub.15701

Krishnan G V, Largay Iii J A, (2000) The predictive ability of direct method cash flow information. Journal of Business Finance & Accounting 27:215-245. https://doi.org/10.1111/1468-5957.00311

Lev B, Li S, Sougiannis T, (2010) The usefulness of accounting estimates for predicting cash flows and earnings. Review of Accounting Studies, 15:779-807. https://doi.org/10.1007/s11142-009-9107-6

22

Mann H, (1984) A Worksheet for Demonstrating the Articulation of Financial Statements. The Accounting Review 59.

Melendrez K D, Schwartz J W C, Trombley M A, (2008) Cash Flow and Accrual Surprises: Persistence and Return Implications. Journal of Accounting, Auditing & Finance 23:573-592. https://doi.org/10.1177/0148558x0802300407

Ozuna T J, Gomez I A, (1994) Estimating a System of Recreation Demand Functions Using a Seemingly Unrelated Poisson Regression Approach. The Review of Economics and Statistics 76:356-360. https://doi.org/10.2307/2109892

Pfeiffer R J, Elgers P T 1(999) Controlling for lagged stock price responses in pricing regressions: An application to the pricing of cash flows and accruals. Journal of Accounting Research 37:239-247. https://doi.org/10.2307/2491406

Revankar N S, (1974) Some finite sample results in the context of two seemingly unrelated regression equations. Journal of the American Statistical Association 69:187-190. https://doi.org/10.2307/2285721

Richardson S A, Sloan R G, Soliman M T, Tuna I, (2005) Accrual reliability, earnings persistence and stock prices. Journal of Accounting and Economics 39, 437-485. https://doi.org/10.1108/ajb-07-2014-0041

Sloan R G, (1996) Do Stock Prices Fully Reflect Information in Accruals and Cash Flows About Future Earnings? Accounting Review 71:289-315

Thomas J K, Zhang H, (2002a) Inventory changes and future returns. Review of Accounting Studies 7:163-187

Thomas J K, Zhang H, (2002b) Inventory Changes and Future Returns. Review of Accounting Studies 7:163-187

Wu C, Xu B, (2008) Deflator selection and generalized linear modelling in market-based regression analyses. Applied Financial Economics 18:1739-1753. https://doi.org/10.1080/09603100701735904

Xu H, Dao M, Wu J, (2018) The effect of local political corruption on earnings quality. Review of Quantitative Finance and Accounting.1-24. https://doi.org/10.1007/s11156-018-0758-x

Zellner A, (1962a) An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. Journal of the American Statistical Association 57:348-368. https://doi.org/10.2307/2281644

Zellner A, (1962b) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American statistical Association 57:348-368. https://doi.org/10.1080/01621459.1962.10480664

Zellner A, (1963) Estimators for seemingly unrelated regression equations: Some exact finite sample results. Journal of the American Statistical Association 58:977-992. https://doi.org/10.2307/2283326

Zellner A, Huang D S (1962) Further properties of efficient estimators for seemingly unrelated regression equations. International Economic Review 3:300-313. https://doi.org/10.2307/2525396